A data-driven framework to select a cost-efficient subset of parameters to qualify sourced materials

Nishan M. Senanayake *[1], Jennifer L.W. Carter [1], Cheryl L. Bowman [2], David L. Ellis [2], Joshua Stuckner [2].

Corresponding author's email : joshua.stuckner@nasa.gov~~nms105@case.edu~~

[1] Case Western Reserve University, Department of Materials Science and Engineering, Cleveland, OH 44106, United States of America

[2] NASA Glenn Research Center, Materials and Structures Division, Cleveland, OH 44135, United States of America

Abstract

The quality of powder processed for manufacturing can be certified by hundreds of different variables. Assessing the impact of all these different metrics on the performance of additively manufactured engineered products is an invaluable, but time intensive specification process. In this work, a comprehensive, generalizable, data-driven framework was implemented to select the optimal powder processing and microstructure variables that are required to predict the target property variables. The framework was demonstrated on a high-dimensional dataset collected from selective laser melted, additively manufactured, Inconel 718. One hundred and twenty-nine powder quality variables including particle morphology, rheology, chemical composition, and build composition, were assessed for their impact on eight microstructural features and sixteen mechanical properties. The importance of each powder and microstructure variable was determined by using statistical analysis and machine learning models. The trained models predicted target mechanical properties with an R^2 value of 0.9 or higher. The results indicate that the desired mechanical properties can be achieved by controlling only a few critical powder properties and without the need for collecting microstructure data. This framework significantly reduces the time and cost of qualifying source materials for production by determining an optimal subset of experiments needed to predict that a given source material will lead to a desired outcome. This general framework can be easily applied to other material systems.

Introduction

Design of manufacturing routes for a material often involves large numbers of cost and time-consuming experiments. It is not always clear which variables are critical to control for achieving the required properties during this iterative manufacturing design. It is impractical to use intuition or conduct randomized trial and error methods to navigate the broad design space in an effort to find which processing and microstructure variables are critical to control in order to achieve target

material properties [1]. Thus, the application of predictive computational tools that provide insight and guidance of the next round of required experiments has become pivotal to accelerate the discovery and realization of new materials [2] [3]. In addition to the traditional computational tools, such as density functional theory (DFT) [4] and CALPHAD based simulations [5], machine learning methods have gained great attention in recent years as a means of gaining insight into optimizing designs in this multidimensional space [1]. Machine learning (ML) methods have been successfully applied to efficiently guide computational DFT or CALPHAD virtual experiments [6], and can be a potential solution for both mining heritage, sparse and un-analyzed data in material databases that are collected from a number of tedious physical experiments, and guiding the next round of physical experiments [7] [8]. There are several reasons for the fewer ML-informed experimental approach success stories in structural materials design [9], such as the relative time needed to collect all the experimental data. The application of a data analytical framework that accounts for engineering constraints in data collection and analysis provides a solution to both reducing data collection time and increasing the ability of engineers to interpret the multidimensional relationships. This provides confidence in the approach to select optimal designs of experiments to qualify materials and manufacturing routes saving time and money in the development of a product.

Statistical analysis and feature selection methods have been previously used for various material design and manufacturing purposes [10-19]. Chen et al. have employed regression analysis to develop models to select the most important machining parameters to predict the target variables of work piece quality, cost, and machining time for nickel alloy [10]. The model predicts the target variables with 90% accuracy. Balachandran et al. explored the trends in the structure, chemistry, and/or bonding for the given chemical composition of the class of materials referred to as apatites to search for new materials using machine learning-based feature selection techniques and showed that the selection of key features significantly accelerated the search of the design space [11]. Yang et al. studied machine learning approaches for screening the key features affecting the hardness of high entropy alloys and achieved the 24.8% increment in hardness based on the predictions [12]. Kalidindi et al. have used two-point correlations and principal component analysis (PCA) to build reliable structure-property linkages for dual-phase steels [13]. Rajan et al. employed PCA on a high-temperature superconductors database that contains about 600 compounds to identify the parameters that govern the important properties [14]. Ramesh et al. studied the influence of output parameters with respect to input parameters for fused deposition modeling for Nylon material by employing the Analysis of Variance (ANOVA) test. The optimized value of tensile strength was predicted as 43.953 ± 1.624 MPa with a 95% confidence interval where the actual corresponding mean value is 43.25 MPa [15]. Ruan et al. proposed a sequential feature selection approach to detect the sulfur and phosphorus in steel with the accuracy of 0.99 R^2 value [16]. Munir et al. compared two different recursive feature elimination methods for feature selection for mechanical properties of a polymer product based on multi-dimensional processing data. They showed that the technique combined with the random forest regression yields the best predictive performance [17]. Rajan et al. employed the partial least squares method, to develop a quantitative structure-activity model to predict bulk modulus of AB_2N_4 spinels with a number of input parameters [18]. Ranking-based feature selection methods such as information gain and Pearson correlation

approaches have been employed by Agrawal et al. to construct predictive models for fatigue strength of steel and achieved the R^2 values over 0.97 [19].

In this work, we developed a data-driven framework that can 1) account for engineering constraints in data collection method, 2) integrated machine learning methods to conduct an unbiased assessment of feature selection and modeling approaches, and 3) determine the optimal set of input variables for prescribed target variables. The utility of the framework was used to demonstrate the reduction in time/cost necessary to qualify the powder feedstock for additively manufactured (AM) Inconel 718 (IN 718) when optimal input features are selected.

IN 718 is an important nickel-based superalloy used extensively in engineering applications such as gas turbines, jet engines, nuclear reactors, pumps, and tools because of its weldability, excellent moderate-temperature mechanical and corrosion-resistant properties [20~~10~~][2~~1~~1]. The manufacturing of IN 718 following conventional manufacturing methods such as casting, forging, and subtractive machining is challenging due to the work-hardening effect and geometrical flexibility [2~~1~~2]. The rapid development of powder bed fusion AM in recent years has led to a great commercial interest in this alternative method for manufacturing this in-demand alloy [2~~1~~3]. IN 718 powder feedstock is produced commercially by both gas and rotary atomization, and the typical powder diameters vary between 10 and 45 µm [2~~1~~4]. Due to a number of concomitant features in the production of commercial feedstock, there is considerable variability among powder lots among powder lot suppliers [21~~11~~]. The microstructure and resulting mechanical properties of the AM IN 718 has been shown to depend on atomization procedures that impact powder rheology, sizes, and chemical composition [2~~1~~5][2~~1~~6]. Quantifying all possible powder lot variables from different vendors requires numerous cost and time-consuming experimental setups [2~~1~~7]. We could specify that all variables should be collected, but this is not an engineered, efficient solution. Therefore, developing a framework for applying ML-enabled models that determine critical experiments needed to verify that a powder lot will likely produce components with desirable mechanical properties is valuable. This generalized framework allows different stakeholders in the material manufacturing development cycle to value different efficiency approaches (i.e., time, cost, uncertainty) for assessing technology risk level. The framework can be adaptable to incorporate new/alternative high-dimensional datasets and modeling approaches as they arise.

To address these challenges, we developed a data-driven framework applicable to both storing and analyzing high-dimensional data, applying different models to enable engineers to make informed decisions when selecting the optimal variables that impact certain target variables. In this work, the data were collected from 18 powder lots of IN 718 feedstock (that include 15 pristine powder lots from five suppliers and 3 recycled powders). In total, 129 processing and 8 microstructure variables were assessed for their statistical correlation to 16 different mechanical properties of additively manufactured IN 718. In order to accommodate the common problem of the propensity

of models to overfit because materials data arises from a small number of observations and high degree of dimensionality, the analysis required carefully selecting data science and machine learning approaches. The objective was to select the minimum number of important processing and microstructural input variables based on their ability to predict the target within a user-defined threshold of uncertainty. The framework trains models to establish relationships (processing-microstructure, microstructure-property, and processing-property) with the optimal/minimum number of input variables. For example, the powder processing and microstructure input variables that are most important to achieve the targeted material properties. First, microstructure models can be compared with the domain knowledge to understand the underlying physical mechanisms. Additionally, the framework provides insight into which variable characterizations are unnecessary in the next design of experiments. This can save significant time and cost by designing an optimal experiment setup with the fewest number of tests necessary to qualify a powder lot. This framework can be used to find a cost-effective subset of variables to track in order to determine whether source materials from different vendors can be substituted and still achieve design requirements.

Methods

Data-Driven Framework

The data-driven framework was established in Python and includes pathways to all the necessary data variables, and is a collection of codes that conduct exploratory data analysis (EDA), various feature selection methods, and test models for a required level of accuracy. It comes with a collection of terms that are worth defining. First, variables are measured experimentally and can be categorized into different metrics. Metrics are the categories of model relationships that are of interest: processing, microstructure, and performance. A model attempts to significantly describe the relationship between a target variable and the other variables by selecting the minimum number of variables that are statistical features of the target. Here we demonstrate the framework by working through models of processing-property and microstructure-property metrics, but it is possible to create models for target variables from a combination of features within any metric (i.e., fatigue strength target variable, a performance metric, can be modeled by a functional relationship with the features of tensile strength and carbon content which are variables in metrics performance and processing respectively).

EDA are data summary and visualization approaches to discover the patterns and understand the data. These approaches often provide insight into variables with low variance or are highly correlated (i.e., particle diameter and particle area). Feature selection methods reduce the overfitting and the complexity of the model, which increases model accuracy and aids interpretation [21~~+~~8]. There are a number of potential supervised feature selection techniques that can be subdivided into three method classes: filter, sequential, and embedded [21~~+~~9]. The filter

methods measure the relevance of variables with statistical measures such as t-tests, ANOVA, and Pearson correlation coefficient [320][321]. These feature selections techniques are useful in removing low variance variables. The sequential methods uses an objective function such as the performance of a classification or regression model to rank-order features [3212]. Finally, embedded feature selection methods conduct the process within the construction of the machine learning algorithm itself and perform selection during the model training [3223]. All three methods were used to select the best predictive features that impact certain targets.

The general procedure is as follows. First, all categorical variables (i.e., non-continuous features with discrete classes) in the dataset were one-hot encoded during the preprocessing step, so that only feature selection and modeling methods dealing with numerical variables were necessary. For each target variable model, two filter methods were applied to the whole set of categorized variables to select and remove the low variance and highly correlated variables. Then the sequential and embedded feature selection methods were applied to the various models. Then the models were compared, and the one with the highest R^2 value is labeled the best. Finally, we check that the model conforms to the assumptions. This is particularly important for regression models which have assumptions of normally distributed independent residuals [3234]. The filter and sequential feature selection method are explained below in separate sections, but the embedded methods are explained in the relevant modeling techniques sections.

The framework was established to demonstrate variance threshold, and Pearson's coefficient filter feature selection techniques, and ridge regression for the sequential feature selection method. Random forest and Extra Tree models were selected to demonstrate embedded feature selection methods. This framework is expandable to other machine learning modeling approaches as domain knowledge grows, or the data set sparseness decreases.

Filter Feature Selection

Remove low variance variables

A variance threshold filter was applied to each numeric variable to select the low variance variables. The variables were normalized to a scale of 0 to 1 so that a single threshold value could be applied. The low variance variables contain the same or almost the same value throughout the whole dataset. Variables with near-zero variance increase the model complexity (i.e., dimensionality) without providing any information about the target. The variance of the variables (σ) was calculated [3245], and normalized features whose variance did not meet the threshold value of 0.001 (relative variance is less than 0.1%) were removed from the dataset.

Remove highly correlated variables

Mutually correlated (multicollinearity) variables convey redundant information, thus incorporating both would again increase model complexity without providing insight about the target. Filtering by the Pearson coefficient is a method to assess the correlation among predictive variables. It calculates linear correlations between pairs of variables and returns a range of values

from +1 to -1. A value of 0 indicates no association between the two features. Positive and negative values indicate proportional and inversely proportional association between the two variables respectively. The Pearson coefficient, r, between two variables, x and y, was calculated using Eq. 1 [30~~19~~].

$$r = \frac{\sum(x_i - \overline{x})\,(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2\ \sum(y_i - \overline{y})^2}} \qquad \text{Eq. 1}$$

Where x_i stands for the value of i^{th} observation and $\overline{x}$ is the mean of variable x respectively, and y_i and $\overline{y}$ are the compared observation and mean of variable y respectively.

Removal of correlated variables was done one way or another if variables x and y were measured simultaneously, or not, by the same experiment. The 129 processing variables were subdivided such that all variables within a group are measured simultaneously by the same experiment. Each group was given a priority number based on the cost-time consumption of the experimental setup; with low priority given to high cost-time consumption. Pearson coefficient was calculated for each pair of variables (independent of group). For highly correlated variables with a correlation coefficient greater than 0.9 or less than -0.9, the variable in the group with the lower priority number was removed until no highly correlated features remained. If the correlated variables were in the same group, a randomly chosen one was removed as there is no difference in the cost-time consumption.

Sequential Feature Selection

Sequential feature selection algorithms are greedy search algorithms that select the most relevant subset of features (k) to the target from the initial d-dimensional feature space where $k < d$ [32~~5~~6]. Sequential feature selection removes or adds one feature at a time to the selected model and is based on the given objective function until it reaches the optimal number of the feature [32~~6~~7]. In this work, ridge regression [32~~7~~8] was selected as the model and the mean squared error [32~~8~~9] was used as the objective function. By convention, the algorithm always tries to minimize the value of the objective function. Therefore, mean squared error that calculated with repeated K-fold cross-validation was used to validate the sequential feature selection results. In this approach, the dataset was split into random (K= #) smaller subsets [39~~30~~]. The model is trained K times using a different subset as the validation set each time while the remaining subsets were used for training. This process is repeated K times with different randomly chosen subsets in each repetition.

Sequential feature selection can be further subdivided into three sub-categories based on the sequence direction, namely Sequential Forward Selection (SFS), Sequential Backward Selection (SBS), and Sequential Forward or Backward Floating Selection (SFFS or SBFS). In SFS, the search for relevant features begins with a model of the target from an empty set, and features are iteratively added until a given number of features are found or the desired model accuracy is

achieved. In SBS, the search begins with a model of the target containing all features in the variable dataset then each are iteratively removed until an optimal number of features are found [40~~18~~]. The SFFS and SBFS are an extension of the SFS and SBS algorithms with an additional exclusion or inclusion step to remove features once they were included [40~~31~~]. In the framework, all sequential feature selection methods were implemented and applied using a representative K-fold value of 3 (i.e., such that each train/test group of data samples is large enough to be statistically representative of the broader dataset).

Ridge regression

Ridge regression was selected as the model to integrate with the sequential feature selection. The integrated estimators such as linear regression used in feature selection may suffer from overfitting when the number of features is large in comparison to the number of observations [37~~27~~]. Ridge regression is an alternative method to ordinary least squares linear regression that penalizes non-zero model features with an additional regularization term added to the loss function (residual sum of squares) that is minimized during feature selection. This is called L_2 regularization [41~~32~~], and the amount of regularization is controlled by a parameter, λ, as shown in the loss function presented in Eq. 2.

$$L = \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda \sum_{j=1}^{p}\beta_j^2 \qquad \text{Eq. 2}$$

Where y_i stands for the target variable of the i^{th} observation, x_{ij} is the j^{th} feature of the i^{th} observation, and βs are the relevant coefficients of the features. In this study, the λ hyperparameter was estimated by cross-validation to give the optimal tradeoff between bias and variance by minimizing the mean squared error of the model on the hold out (validation) sets. Hyperparameter optimization was performed using the random search algorithm that is implemented in the Scikit-learn Python library [42~~33~~]. Random search sets up a grid of hyperparameter values and selects random combinations to train the model and to discover the best hyperparameter.

Model assumptions for ridge regression

For a ridge regression model of a target variable from the features in the data set to be valid, four assumptions must be met [3~~3~~3]. First, a linear relationship exists between the features and the target variable. If the relationship between features and the target variable is not linear then the analysis might misinterpret the true relationship. This is most easily checked by visualizing predictions vs residuals plots

Second, is homoscedasticity that the residuals (difference between the observed and estimated values) are similar (constant variance) across the values of the independent variables. When the variance of residuals differs at different feature values (i.e., high/low Carbon content),

homoscedasticity is violated. The assumption is checked by the Breusch-Pagan test and the visual examination of a plot of predicted values vs. residuals [4334]. A calculated p-value from the Breusch-Pagan test of less than 0.05 indicates homoscedasticity.

Third, there is no correlation between consecutive residuals, meaning that the residuals are independent. The assumption was checked by using the Durbin-Watson Test [4435]. The Durbin-Watson (d) value can range from 0 to 4. Calculated d values between 1.5 and 2.5 indicate there are no serious autocorrelation in the residuals.

Fourth, and final, is that the residuals are normally distributed. This assumption is checked quantitatively by using the Anderson-Darling test [4536] and generating the quantile-quantile (Q-Q) plots for each prediction. If the points on the Q-Q plot approximately form a straight diagonal line, then the normality assumption is met for the model.

Embedded Feature Selection with Random Forest and Extra Tree Regressor

Random Forest (RF) and Extra Tree (ET) regressors are two ensemble models that have many features in common. Both regressors are composed of a large number of decision trees, where the final decision is obtained taking into account the prediction of every tree [4637][4738]. Feature selection with RF and ET are embedded methods within the construction of the machine learning model. A subset of features is selected at each node and the best split is chosen based on the performance. In feature selection, the algorithm ranks the importance of the features based on the order of features that minimizes the Gini impurity [4637]. The value of the Gini impurity represents how good the split is at each node. There are two main differences between the two techniques. RF uses bootstrap replicates, which means that the same data points can be selected multiple times to build an individual tree while ET uses the entire original sample set. Additionally, RF chooses the optimum split while ET chooses it randomly. However, once the split points are selected, the two algorithms choose the best one from all the subsets of features. Therefore, ET adds randomization but still has optimization. These differences make ET computationally more efficient. Both techniques have shown to be potential candidates for high-dimensional data and perform well in the presence of noise due to their predictive performance and low overfitting [4738].

The models' hyperparameters were tuned by using random search using the RandomizedSearchCV module from the scikit-learn library [4839]. Hyperparameters were randomly sampled from a uniform distribution of possible parameter values and the selected hyperparameter was set for each regressor that gave the best accuracy for predicting the target.

Model efficacy

The efficacy of the different models (ridge, RF, and ET) implementing different feature selection methods was estimated by calculating adjusted R^2 [490]. Adjusted R^2 value can range from 0 to

1, where the value 1 indicates that the model predictions fit perfectly with the data. It represents the proportion of variance (of the target) that has been explained by the features in the model.

Data Description

Dataset was collected at NASA Glenn Research Center (Cleveland, OH) and contains measurements of variables from 18 powder lots: 15 pristine and 3 recycled (i.e., used for additive manufacturing a component already) which were produced by either gas or rotary atomization in either Nitrogen or Argon atmosphere. The data types are both categorical (i.e., non-continuous features such as grain structure with discrete recrystallized, partially recrystallized, and anisotropic classes and numerical (e.g., grain size). The dataset includes measurements from chemical, metallographic, particle size, morphology, flow, rheology, and mechanical testing. The cost-time consumption weighting parameter for each experiment is presented in Table 1. Post processing heat treatment was sufficiently controlled across all powders and it had a negligible detrimental effect on samples. The values of weighting parameters can be changed based on different efficiency approaches (i.e., time, cost, uncertainty) for assessing technology risk level. A second iteration design of experiments that reduce the number of low priority instrument could greatly increase the powder qualification rate The full description of the experimental methods is included in Sudbrack et al., [2~~1~~4] and each methodology is explained briefly in this section.

Table 1. Lists the cost-time priority value for the fifteen different instruments that are needed to collect the 156 variables from the three different metrics (processing, microstructure, and performance). Metrics groups are indicated by shading.

Data	Instrument	Metric	Number of Variables	Cost-Time Priority
Particle size	Malvern Morphologi G3SE	Processing	77	1
Rheology data	Rheometer	Processing	17	2
Powder C/S content	LECO CS-444-LS	Processing	2	3
Powder N/O content	TC-436 N/O	Processing	2	3
Powder B+M content	Varian Vista Pro Inductively Coupled Plasma (ICP) Emission Spectrometer	Processing	14	3
Build C/S content	LECO CS-444-LS	Microstructure	2	4
Powder N/O content	TC-436 N/O	Microstructure	2	4
Build B+M content	Varian Vista Pro Inductively Coupled Plasma (ICP) Emission Spectrometer	Microstructure	13	4
Porosity	Optical Microscopy	Microstructure	2	5
Nitrides and Carbides	SEM	Microstructure	4	6
Grain size	EBSD	Microstructure	2	7
Surface Roughness	Roughness	Performance	1	8
Hardness	Vickers Hardness	Performance	1	9
Tensile data	Tensile	Performance	6	10
Reduction in area	Calipers/micrometer	Performance	2	11

Fatigue data	HCF	Performance	6	12

Processing metrics

Powder size distribution: Malvern Morphologi G3SE system and Horiba LA-950V2 were used to collect the powder size data. A minimum of 20,000 individual powder particles (per scan) were imaged using transmitted illumination and automated optical image collection, scan area with 25 mm diameter, and minimum feature size of 30 pixels. The powder particles with the value particle area divided by a convex hull area of less than 0.9 were excluded. The morphology data includes 77 different variables for the circularity, aspect ratio, elongation, and convexity of the particles. Number-based and volume-based size distribution were also measured using laser diffraction with Horiba LA-950V2 analyzer.

Powder chemistry: The content of the 18 elements in the powder chemistry were quantified from the average of two measurements from three different instruments: a LECO CS-444-LS carbon/sulfur determinator, a LECO TC-436 nitrogen/oxygen determinator, and a Varian Vista Pro Inductively Coupled Plasma (ICP) Emission Spectrometer for the remaining boron and the metallic elements.

Rheology: was performed using the Freeman Technology FT4 powder rheometer. Approximately 200 grams of powder of each type was used and duplicate runs were performed. The powders were characterized by important rheology features including basic flow energy, stability index, flow rate index, specific energy, aeration ratio, normalized aeration sensitivity, pressure drop, permeability, consolidation energy, consolidation index, and compressibility.

Microstructural and Performance Metrics

Coupon samples: of bulk materials were produced for characterization on a Concept Laser M1 selective laser melting machine equipped with a custom-built box (100 mm × 100 mm × 80 mm) using the manufacturer supplied standard build parameters for IN 718. The characterization of microstructure and performance metrics was done at two thermal stages: green state (no thermal post-processing) and heat-treated.

Build chemistry: The build chemistry data is technically a microstructure metric since it arises after additive manufacturing, but for this demonstration there should be near perfect correlation between powder and build chemistry. Therefore, comparing these variables with the powder chemistry variables using filter feature selection techniques is the only analysis possible at this time.

Microstructure: Optical microscopy was used to measure porosity size porosity volume fraction, grain size, and grain structure. Scanning electron microscopy (SEM) was used to measure the size and area fractions of secondary phases: nitrides were quantified with secondary electron micrographs, and carbides were quantified from backscattered electron micrographs. The grain structure of the specimens was quantified using electron backscatter diffraction (EBSD).

Performance testing: Standard mechanical testing was conducted including tensile testing which measures variables of elastic modulus, proportional limit, yield strength (0.02 and 0.2 offset), tensile strength, elongation to failure, and post-fracture analysis of reduction in cross-sectional area. High cycle fatigue (HCF), which collected the variables such as average HCF life, max HCF life, minimum HCF life, average HCF stress, max HCF stress, and minimum HCF stress. The surface roughness was characterized by a high precision profilometer (Tylor Hobson's Form Talysurf PGI 1200), which is equipped with a 2 μm radius diamond canonical sphere tip. Measurements were made along with both longitudinal and transverse directions. The Hardness measurements were conducted using a Struers Durascan with the Vickers tip and the average of twenty measurements was included in the final dataset.

Results and Discussion

Low variance features

The calculated variable variance indicates the data for 16 processing variables related to the particle size distribution (13) and chemical analysis (S, O in the powder and build) had a variance lower than the threshold value of 0.001. It can be concluded that the impact of these 17 variables on the targets is not explicated and they were not further analyzed, reducing the dimensionality complexity of the processing variable space from 129 variables to 112 variables. In the case of the particle size distribution variables, this could indicate that it is unnecessary to calculate these variables during analysis.

The lack of variance in the elemental variable content motivates future research and development efforts, as additional data with a larger variance is required to understand how these variables impact performance. For example, there is insufficient variance in oxygen content in this dataset to model the impact of this variable on other target variables, but control and removal of oxide inclusions is a foundational challenge in superalloy production [5041]. Additionally, facility with access to different instruments, or value added weighting could make an alternative decision. If a facility has access to an XRF, than measuring build chemistry (including oxygen content) might be a preferred quality control approach since under optimal AM conditions the oxygen can segregate to the spatter and not the build [5142]. Alternatively, if the supplier is interested in reusing powder then monitoring oxygen content in the powder makes sense.

Highly correlated features

Correlations among variables, mainly particle size data, powder and build chemistry, and rheology were visualized using the correlation matrix as a heat map (powder|build chemistry shown in Fig. 1). This step of exploratory data analysis (EDA) provides insight into the correlation between individual features. Fig. 1 shows the correlation heat map for powder and build chemistry features.

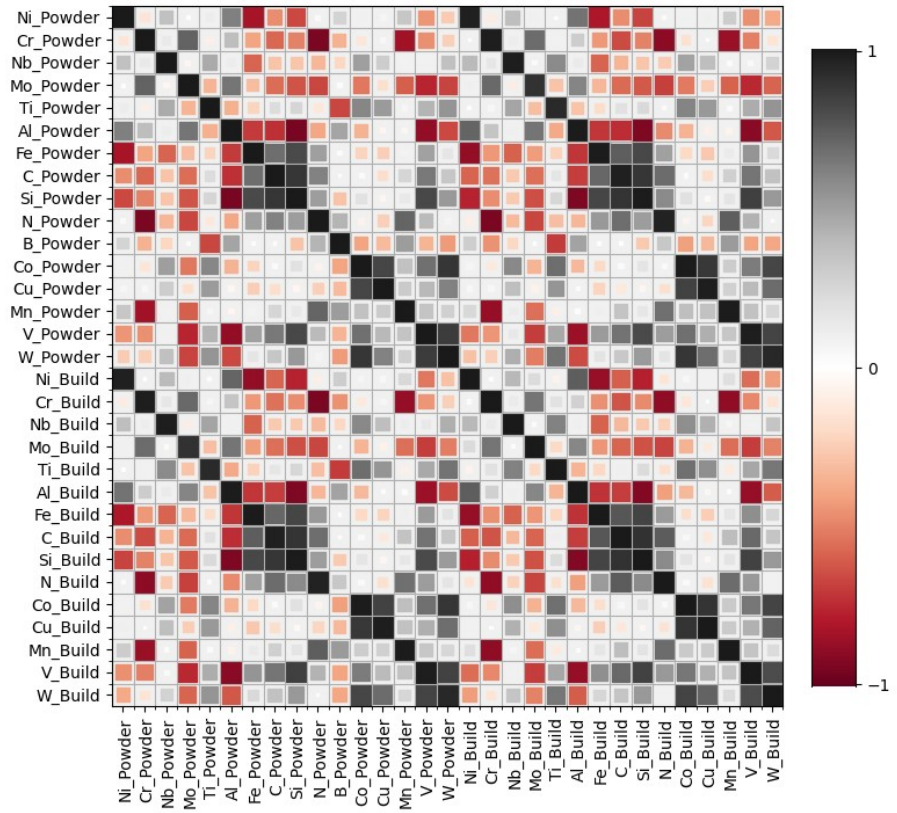

Fig. 1. The Pearson coefficient heat map for powder and build chemistry features. Highly corelated features are denoted with relatively larger size of the squares in dark gray (positively corelated) and dark red (negatively correlated).

The gradient of colors and sizes represents the intensity of correlation of each pair of features. The relatively large and more saturated red and grey colors represent increasing proportional and inversely proportional correlation, respectively. White represents no correlation between two features. As an example, Nickel composition of the powder (Ni_powder) and the build sample (Ni_build) are near perfectly proportional, as anticipated. In Table 1, the cost-time penalty for measuring build chemistries is higher than the powder chemistries. Thus, the Ni_build variable could be excluded from the data set as the group that includes Ni_build (build_group) has been assigned a lower priority than the group that includes Ni_powder (powder_group).

The Pearson coefficient calculations for the exhaustive set of all feature pairs indicates 58 redundant features which are subsequently removed from further analysis based on their priority

number. In all cases, the redundant features were correlated (Pearson coefficient > |0.9|). These redundant features include 41 particle size distribution variables, and the 15 build chemistry and 2 powder chemistry variables. The remaining potential features include 23 particle size distribution variables, all the rheology variables, and the 14 powder chemistry variables. The combined filter feature selection techniques reduce the dimensionality complexity from 129 variables to 54 variables for further modeling.

Insight from the dimensionality reduction can be used to narrow down the design space for future experiments. The dataset represents the first design of experiments which required collection of, ultimately, redundant features, such as Ni_build which correlated with Ni_powder. For future research and development efforts, this second measurement of Ni_build could be avoided to save time and money. According to this analysis, the chemistry of the build samples can be excluded when qualifying future powder lots as the build chemistry cost-time priority less than the powder chemistry (shown in the Table 1). Alternatively, in quality control process, measuring redundant variables, or retaining samples, could prove beneficial for troubleshooting processing errors. For example, a lack of correlation between the nickel concentration in the build and the powder could indicate that a machine was not properly cleaned between different powder compositions.

This EDA also can provide insights into the methods that different vendors choose to control chemistry. The highly inversely proportional (dark red boxes) elude to the robust history of modifying chemistry in superalloys [52]~~-. [43].~~ This particular dataset is insufficient for in-depth screening of the alloy design space since it only contains 18 different powder lots. A more diverse dataset, like the one compiled for the family of 9-12wt% Cr steels, could be used to guide researchers on the historical ~~[44]~~ and new alloy design ~~[45]~~ choices to advance alloy performance [53][54]. These examples are provided to highlight the selection and design of experiments is dependent on the objective of the data and no one priority method will be suitable for all. This complexity motivates the need for a framework so that different EDA, and priority schemes can be evaluated.

Process - Structure Model

After the filter feature selection steps, which removed low variance (17) and highly correlated (58) variables, ridge regression modeling with sequential feature selection and RF and ET models with embedded feature techniques were applied to model microstructure target variables (8 total) from features in the remaining set of Processing variables (54 total). For all microstructural target variables, Sequential Forward Floating Selection (SFFS) accompanied with ridge regression produced the process-microstructure models with the best accuracy and least number of features. All the R^2 values for the models' predictions are listed in the Table 2.

Table 2. R^2 values of predictions for all the microstructural target variables

Target Variable	**Number of features**	**R^2**
Porosity size	4	0.95
Porosity volume fraction	4	0.96
Grain size	5	0.91
Grain structure	6	0.89
Average diameter of carbides	4	0.89
Volume fraction of carbide	4	0.91
Average diameter of nitrides	5	0.88
Volume fraction of nitrides	4	0.89

In this section, the results are demonstrated for one microstructure target variable: porosity size. The Ridge regression model with the lowest standard error, selected the most important features, illustrated in Fig. 2. Fig. 2a shows the variance of the standard error with the number of the features and the error approaches 0.1 when the model includes the four most impactful features. The uncertainties, calculated by using the repeated K-fold cross-validation, as shown in the light blue shade in Fig. 2a, decrease with added model complexity. The correlation and the importance of each feature generated using ridge regression are presented in Fig. 2b. The higher coefficient magnitude indicates the larger impact on the target, so in ranked order of importance: skewness of the particle powder size distribution, consolidation bulk density, standard deviation of powder size distribution and molybdenum content in the powder for 95% (corresponding R^2 value) of the variance in the target variable of porosity size. Predicting that the particle size variables such as skewness and standard deviation of the size distribution has the highest impact on the porosity size is consistent with the many previous studies that indicated the porosity size varies with the size metrics of the powder particles in additive manufacturing [55~~46~~] [56~~47~~]. Additionally, the predictions show that the porosity size is influenced by the conditioned bulk density of the powder. Conditioned bulk density is determined by the particle shape and particle size distribution and it is known that higher powder density creates dense components with smaller porosities [57~~48~~]. The effect of Mo composition on the porosity size has not been explored to date in the literature and represents additional potential tuning parameters that process engineers can specify for powder to meet specific porosity quality targets.

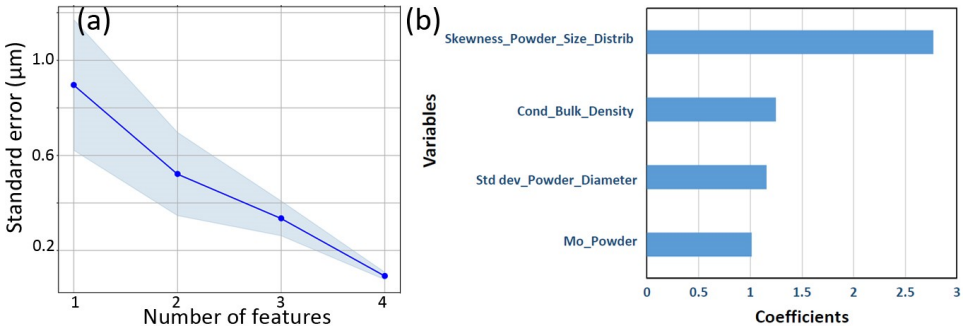

Fig. 2(a) shows the standard error against the number of features (b) represents the absolute values of coefficients of selected features to the target variable (porosity size).

The regression model for porosity volume size has a calculated R^2 for the fitted regression line of 0.95, and the prediction plot is shown in Fig. 3a. Visual interpretations of the results of linearity and homoscedasticity (Fig. 3b), and normal distribution of the residuals (Fig. 3c) indicate that the predicted values are homogeneous over the entire range of actual values and conform to the assumptions necessary for a valid linear regression.

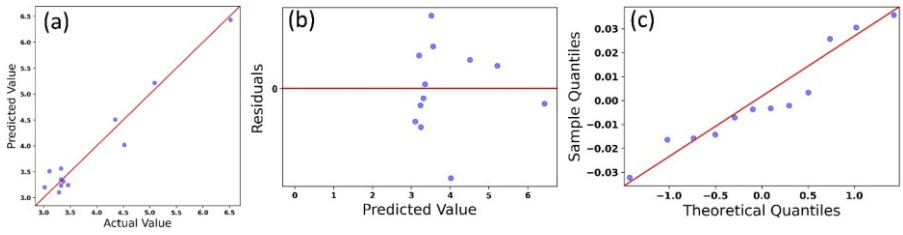

Fig 3. (a) Shows the comparison of measured versus predicted porosity volume fraction. (b) shows predicted porosity volume fraction versus residuals that were presented to check the linearity and homoscedasticity. (c) shows the quantile-quantile plot that was presented to check the normality of the residuals.

As shown in Fig. 3b, no significant curvilinear relationship is existing between residuals and predicted values. The residuals are randomly scattered around 0 (the horizontal line) providing a relatively even distribution (shown in Fig. 3b). The calculated p-value from the Breusch-Pagan test was 0.04 which fails to reject the null hypothesis for the presence of homoscedasticity. Thus, it can be concluded that the ridge model for porosity volume fraction meets the assumption of homoscedasticity. The Durbin-Watson value for the model was calculated to be 2.1 which is in the range of 1.5 and 2.5 that is considered to have no significant autocorrelation in the residuals (i.e., third assumption). The fourth assumption (i.e., normality), confirmed visually in the generated quantile-quantile plots shown in Fig. 3c, was confirmed by the Anderson-Darling test which

confirmed that the residuals are normally distributed with a significance level of 0.1. Therefore, we can assume that our model also meets the normality assumption.

Structure – Mechanical Property Model

In this section, the sixteen mechanical properties were selected as the target variables, and models selected potential features from the eight microstructural variables. Only one target, 0.2% yield strength, was explained with an R^2 of 0.9. The four most important features for the random forest regressor model of 0.2% yield strength are generated in the map of importance shown in Fig. 4a (using the feature_importances module in scikit-learn) [58~~49~~]. It shows that the porosity size, volume fraction of the carbide and nitride, and the grain size have the most significant impact on the yield strength. The predicted values for the random forest regressor and the measured values are compared in Fig. 4b.

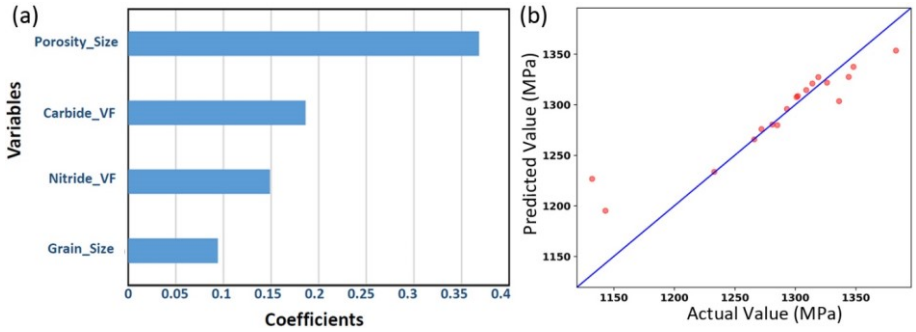

Fig. 4 (a) Shows the absolute value of rank-order importance of different features and (b) shows the comparison of measured versus predicted 0.2% offset yield strength by the random forest regressor.

As expected, the yield strength is influenced by the porosity size, volume fraction of carbide and nitride, and grain size. Porosity has an obvious negative effect on yield strength for all materials. It is known that the MC and $M_{23}C_6$ carbides at grain boundaries play an important role in affecting mechanical properties which are detrimental to both tensile strength and creep resistance in nickel-based superalloys [59~~0~~]. Antonov et al., has shown that nitrides have a mixed effect on the mechanical properties depending on the type of compound that is formed [60~~51~~]. Mangen et al., studied that the grain size effect on the yield strength by following the Hall-Petch relation, and the slope of the Hall-Petch relation is proportional to the spacing between precipitates [61~~52~~]. Thus, it can be concluded that the results from the data science framework are well-aligned with the conclusions from the previous studies.

Process - Mechanical Property Model

As in the previous section, the mechanical property variable of 0.2% offset yield strength was set as the target, but the processing variables rather than microstructural variables were used to compare different modeling methods. Among the applied techniques, the model with the highest R^2 value (0.97) for 0.2% offset yield strength was given by SFFS and ridge regression. Fig. 5a shows the change in standard error with the number of features. The relative error that compares to the actual values becomes less than 90 MPa when the model includes the five most impactful features. The uncertainties shown in the light blue shade in Fig. 5a, calculated by using the repeated K-fold cross-validation, decrease with increasing number of features and level off to ± 30 MPa with five features, thus indicating diminishing impact of additional features. The five most important features that affect the 0.2% offset yield strength and their coefficient values are shown in Fig. 5b.

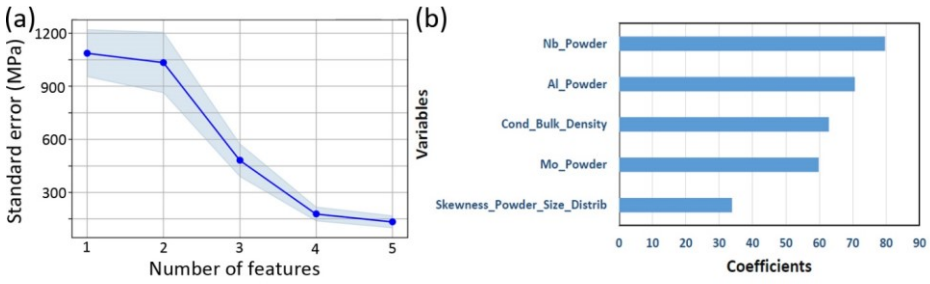

Fig. 5(a) shows the standard error against the number of features and 5(b) represents the absolute values of coefficients of selected features to the target variable (0.2% yield strength)

Obviously, niobium and aluminum content in the powder should positively impact yield strength, since the formation of γ″ (Ni_3Nb) and γ′ (Ni_3Al) precipitates are the primary strengthening mechanisms in IN718[10]. The effect of consolidation bulk density, Mo composition and skewness of the powder size distribution can be hypothesized as an indirect dependence of porosity sizes that is shown in Fig. 2b in process-structure section [62~~53~~]. Mo is a known solid solution strengthener of the Ni matrix that effects on pores as a beneficial side effect [63]~~54~~]. The same statistical tests were also conducted for the processing-structure models to confirm the assumptions of multiple linear regression.

The developed framework narrows down the design space that needs to be controlled to achieve the desired material properties. The insights can be used by researchers in various ways. First, insights could be utilized to skip uninformative or redundant experiments in the next round of optimization experiment to increase the rate of research and development. As an example, when qualifying the future powder lot for the desired mechanical property such as 0.2% yield strength,

it is adequate to characterize the composition of the powder, skewness of the powder distribution and the consolidation bulk energy. Secondly, insights can be utilized to purposefully select additional experiments to reduce risk during quality control processes. For example, the skewness of the powder size distribution, consolidation bulk density and Mo composition correlate with porosity size. But other variables including grain size, volume fraction of carbides and nitrides are critical microstructure variables that control the yield strength. Therefore, in the quality control process it might be advantageous to continue to specify and measure carbon and nitrogen content of the powder. Process-structure models can be used to predict carbide and nitride fraction instead of measuring these parameters because of the time/money associated with the data collection process.

Conclusion

In this work, we have shown that a small subset of powder, build sample, and microstructure variables can be used to accurately predict target properties of the AM IN 718. Several feature selection and modeling techniques were integrated into a generalizable framework to select the best models that can accurately explain the desired targets with an optimal cost-efficient subset of features. Due to high dimensionality of the dataset and the small number of observations, robust cross-validation was used to limit overfitting and confirm that relevant statistical assumptions were not violated. The results from the process-structure models will help researchers to understand the underlying metallurgical principles involved in additively manufactured IN 718. The structure-property models help to determine which microstructural features are most impactful in producing the desired material property. The process-property models and feature selection allow researchers and manufacturers to determine an optimal and cost-efficient subset of features that should be controlled to qualify that sourced ingredients will likely lead to desired product outcomes.

Data and Code Availability

Formatted: Heading-2

The code and data required to reproduce these findings is available to download from https://github.com/nasa/feature_selection_tool and/or from the corresponding author upon request.

Acknowledgments

The contributions of the authors from Case Western Reserve University (Senanayake and Carter) was paid for by NSF Non-Academic Research Internships for Graduate Students (INTERN) Supplemental Funding Opportunity (DCL: NSF 18-102) to NSF Award (# 1552716) CAREER: A Systematic Data-Analytics Approach to the Design of Interface-Rich Materials.

This work was supported by the NASA Transformational Tools and Technologies (T^3) project under the Transformative Aeronautics Concept Program within the Aeronautics Research Mission Directorate.

The dataset was developed for the Additive Manufacturing Materials Integrity Initiative program. The authors would like to thank Brad Lerch (retired) for leading the mechanical testing effort and Chantal Sudbrack for leading the SEM, chemistry, and other data collection.

References

[1] Balachandran, P. V., Xue, D., Theiler, J., Hogden, J., Gubernatis, J. E., & Lookman, T. (2018). Importance of feature selection in machine learning and adaptive design for materials. In Materials discovery and design (pp. 59-79). Springer, Cham.

[2] de Pablo, J. J., Jones, B., Kovacs, C. L., Ozolins, V., & Ramirez, A. P. (2014). The materials genome initiative, the interplay of experiment, theory and computation. Current Opinion in Solid State and Materials Science, 18(2), 99-117.

[3] Senanayake, N. M., & Carter, J. L. (2020). Computer Vision Approaches for Segmentation of Nanoscale Precipitates in Nickel-Based Superalloy IN718. Integrating Materials and Manufacturing Innovation, 9(4), 446-458.

[4] Kohn, W., & Sham, L. J. (1965). Self-consistent equations including exchange and correlation effects. Physical review, 140(4A), A1133.

[5] Kattner, U. R. (2016). The Calphad method and its role in material and process development. Tecnologia em metalurgia, materiais e mineracao, 13(1), 3.

[6] Lookman, T., Balachandran, P. V., Xue, D., Hogden, J., & Theiler, J. (2017). Statistical inference and adaptive design for materials discovery. Current Opinion in Solid State and Materials Science, 21(3), 121-128.

[7] Sargent, P. (1992). Data quality in materials information systems. Computer-aided design, 24(9), 477-490.

[8] Carter, J. L., Verma, A. K., & Senanayake, N. M. (2020). Harnessing legacy data to educate data-enabled structural materials engineers. MRS Advances, 5(7), 319-327.

[9] Carter, J. L., & Verma, A. K. (2021). Informatics-Enabled Design of Structural Materials. JOM, 1-3.

[10] Chen, S. H., & Hsu, C. H. (2021). Using uniform design and regression methodology of turning parameters study of nickel alloy. The International Journal of Advanced Manufacturing Technology, 116(11), 3795-3808.

[11] Balachandran, P. V., Xue, D., Theiler, J., Hogden, J., Gubernatis, J. E., & Lookman, T. (2018). Importance of feature selection in machine learning and adaptive design for materials. In Materials discovery and design (pp. 59-79). Springer, Cham.

[12] Yang, C., Ren, C., Jia, Y., Wang, G., Li, M., & Lu, W. (2022). A machine learning-based alloy design system to facilitate the rational design of high entropy alloys with enhanced hardness. Acta Materialia, 222, 117431.

[13] Khosravani, A., Cecen, A., & Kalidindi, S. R. (2017). Development of high throughput assays for establishing process-structure-property linkages in multiphase polycrystalline metals: Application to dual-phase steels. Acta Materialia, 123, 55-69.

[14] Rajan, K., Suh, C., & Mendez, P. F. (2009). Principal component analysis and dimensional analysis as materials informatics tools to reduce dimensionality in materials science and engineering. Statistical Analysis and Data Mining: The ASA Data Science Journal, 1(6), 361-371.
[15] Ramesh, M., & Panneerselvam, K. (2021). Mechanical investigation and optimization of parameter selection for Nylon material processed by FDM. Materials Today: Proceedings, 46, 9303-9307.

[160] Ruan, F., Qi, J., Yan, C., Tang, H., Zhang, T., & Li, H. (2017). Quantitative detection of harmful elements in alloy steel by LIBS technique and sequential backward selection-random forest (SBS-RF). Journal of analytical atomic spectrometry, 32(11), 2194-2199.
[17] Munir, N., Mulrennan, K., & McAfee, M. (2021, June). Comparison of data summarization and feature selection techniques for in-process spectral data. In 2021 32nd Irish Signals and Systems Conference (ISSC) (pp. 1-7). IEEE.
[18] Suh, C., & Rajan, K. (2005). Virtual screening and QSAR formulations for crystal chemistry. QSAR & Combinatorial Science, 24(1), 114-119.
[19] Agrawal, A., Deshpande, P. D., Cecen, A., Basavarsu, G. P., Choudhary, A. N., & Kalidindi, S. R. (2014). Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters. Integrating Materials and Manufacturing Innovation, 3(1), 90-108.
~~Amato, K. N., Gaytan, S. M., Murr, L. E., Martinez, E., Shindo, P. W., Hernandez, J., ... & Medina, F. (2012). Microstructures and mechanical behavior of Inconel 718 fabricated by selective laser melting. Acta Materialia, 60(5), 2229-2239.~~

[2011] Smith, T. M., Senanayake, N. M., Sudbrack, C. K., Bonacuse, P., Rogers, R. B., Chao, P., & Carter, J. (2019). Characterization of nanoscale precipitates in superalloy 718 using high resolution SEM imaging. Materials Characterization, 148, 178-187.
[2112] Nguyen, Q. B., Nai, M. L. S., Zhu, Z., Sun, C. N., Wei, J., & Zhou, W. (2017). Characteristics of inconel powders for powder-bed additive manufacturing. Engineering, 3(5), 695-700.
[2213] Jiménez, A., Bidare, P., Hassanin, H., Tarlochan, F., Dimov, S., & Essa, K. (2021). Powder-based laser hybrid additive manufacturing of metals: A review. The International Journal of Advanced Manufacturing Technology, 1-34.
[2314] Sudbrack, C. K., Lerch, B. A., Smith, T. M., Locci, I. E., Ellis, D. L., Thompson, A. C., & Richards, B. (2018). Impact of powder variability on the microstructure and mechanical behavior of selective laser melted alloy 718. In Proceedings of the 9th International Symposium on Superalloy 718 & Derivatives: Energy, Aerospace, and Industrial Applications (pp. 89-113). Springer, Cham.
[2415] Jiang, R., Mostafaei, A., Wu, Z., Choi, A., Guan, P. W., Chmielus, M., & Rollett, A. D. (2020). Effect of heat treatment on microstructural evolution and hardness homogeneity in laser powder bed fusion of alloy 718. Additive Manufacturing, 35, 101282.
[2516] Poorganji, B., Ott, E., Kelkar, R., Wessman, A., & Jamshidinia, M. (2020). Materials Ecosystem for Additive Manufacturing Powder Bed Fusion Processes. JOM, 72(1), 561-576.
[2617] Minet, K., Saharan, A., Loesser, A., & Raitanen, N. (2019). Superalloys, powders, process monitoring in additive manufacturing. In Additive Manufacturing for the Aerospace Industry (pp. 163-185). Elsevier.

Formatted: Normal

[27~~18~~] Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. Computers & Electrical Engineering, 40(1), 16-28.
[28~~19~~] Stańczyk, U. (2015). Feature evaluation by filter, wrapper, and embedded approaches. In Feature Selection for Data and Pattern Recognition (pp. 29-44). Springer, Berlin, Heidelberg.
[29~~0~~] Sánchez-Maroño, N., Alonso-Betanzos, A., & Tombilla-Sanromán, M. (2007, December). Filter methods for feature selection–a comparative study. In International Conference on Intelligent Data Engineering and Automated Learning (pp. 178-187). Springer, Berlin, Heidelberg.
[30~~21~~] Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. In Noise reduction in speech processing (pp. 1-4). Springer, Berlin, Heidelberg.
[31~~22~~] Shardlow, M. (2016). An analysis of feature selection techniques. The University of Manchester, 1(2016), 1-7.
[32~~23~~] Lal, T. N., Chapelle, O., Weston, J., & Elisseeff, A. (2006). Embedded methods. In Feature extraction (pp. 137-165). Springer, Berlin, Heidelberg.
[33~~24~~] Osborne, J. W., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. Practical assessment, research, and evaluation, 8(1), 2.
[34~~25~~] Snedecor, G. W. (1934). Calculation and interpretation of analysis of variance and covariance.
[35~~26~~] Pudil, P., Novovičová, J., & Kittler, J. (1994). Floating search methods in feature selection. Pattern recognition letters, 15(11), 1119-1125.
[36~~27~~] Reunanen, J. (2003). Overfitting in making comparisons between variable selection methods. Journal of Machine Learning Research, 3(Mar), 1371-1382.
[37~~28~~] Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67.
[38~~29~~] Wang, Z., & Bovik, A. C. (2009). Mean squared error: Love it or leave it? A new look at signal fidelity measures. IEEE signal processing magazine, 26(1), 98-117.
[39~~0~~] Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. Encyclopedia of database systems, 5, 532-538.
[40~~31~~] Somol, P., Pudil, P., Novovičová, J., & Paclık, P. (1999). Adaptive floating search methods in feature selection. Pattern recognition letters, 20(11-13), 1157-1163.
[41~~32~~] Verducci, J. S. (2007). Prediction and Discovery: AMS-IMS-SIAM Joint Summer Research Conference, Machine and Statistical Learning: Prediction and Discovery, June 25-29, 2006, Snowbird, Utah (Vol. 443). American Mathematical Soc. Accessed: Oct. 11, 2021.
[42~~33~~] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.
[43~~34~~] Cook, R. D., & Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. Biometrika, 70(1), 1-10.
[44~~35~~] Savin, N. E., & White, K. J. (1977). The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors. Econometrica: Journal of the Econometric Society, 1989-1996.
[45~~36~~] Nelson, L. S. (1998). The Anderson-Darling test for normality. Journal of Quality Technology, 30(3), 298. (accessed Oct. 11, 2021).
[46~~37~~] Biau, G., & Scornet, E. (2016). A random forest guided tour. Test, 25(2), 197-227.
[47~~38~~] Vasant, P., Zelinka, I., & Weber, G. W. (Eds.). (2018). Intelligent computing & optimization (Vol. 866). Springer. (accessed Oct. 11, 2021).

[4839] Paper, D., & Paper, D. (2020). Scikit-Learn Classifier Tuning from Simple Training Sets. Hands-on Scikit-Learn for Machine Learning Applications: Data Science Fundamentals with Python, 137-163.

[4940] Cameron, A. C., & Windmeijer, F. A. (1997). An R-squared measure of goodness of fit for some common nonlinear regression models. Journal of econometrics, 77(2), 329-342.

[5041] Chen, X. C., Shi, C. B., Guo, H. J., Wang, F., Ren, H., & Feng, D. (2012). Investigation of oxide inclusions and primary carbonitrides in Inconel 718 superalloy refined through electroslag remelting process. Metallurgical and Materials Transactions B, 43(6), 1596-1607.

[5142] Pauzon, C., Raza, A., Hryha, E., & Forêt, P. (2021). Oxygen balance during laser powder bed fusion of Alloy 718. Materials & Design, 201, 109511.

[52] Paulonis, D. F., & Schirra, J. J. (2001). Alloy 718 at Pratt & Whitney–Historical perspective and future challenges. Superalloys, 718(706), 13-23.

[53] Verma, A. K., Hawk, J. A., Bruckman, L. S., French, R. H., Romanov, V., & Carter, J. L. (2019). Mapping multivariate influence of alloying elements on creep behavior for design of new martensitic steels. Metallurgical and Materials Transactions A, 50(7), 3106-3120.

[54] Verma, A. K., Huang, W. H., Hawk, J. A., Bruckman, L. S., French, R. H., Romanov, V., & Carter, J. L. (2019). Screening of heritage data for improving toughness of creep-resistant martensitic steels. Materials Science and Engineering: A, 763, 138142.

[5543] Farzadfar, S. A., Murtagh, M. J., & Venugopal, N. (2020). Impact of IN718 bimodal powder size distribution on the performance and productivity of laser powder bed fusion additive manufacturing process. Powder Technology, 375, 60-80.

[5644] Rausch, A. M., Markl, M., & Körner, C. (2019). Predictive simulation of process windows for powder bed fusion additive manufacturing: Influence of the powder size distribution. Computers & Mathematics with Applications, 78(7), 2351-2359.

[57] Wimler, D., Kardos, S., Lindemann, J., Clemens, H., & Mayer, S. (2018). Aspects of powder characterization for additive manufacturing. Practical Metallography, 55(9), 620-636.

~~[45] Paulonis, D. F., & Schirra, J. J. (2001). Alloy 718 at Pratt & Whitney–Historical perspective and future challenges. Superalloys, 718(706), 13-23.~~

~~[46] Wimler, D., Kardos, S., Lindemann, J., Clemens, H., & Mayer, S. (2018). Aspects of powder characterization for additive manufacturing. Practical Metallography, 55(9), 620-636.~~

~~[47] Verma, A. K., Hawk, J. A., Bruckman, L. S., French, R. H., Romanov, V., & Carter, J. L. (2019). Mapping multivariate influence of alloying elements on creep behavior for design of new martensitic steels. Metallurgical and Materials Transactions A, 50(7), 3106-3120.~~

~~[48] Verma, A. K., Huang, W. H., Hawk, J. A., Bruckman, L. S., French, R. H., Romanov, V., & Carter, J. L. (2019). Screening of heritage data for improving toughness of creep-resistant martensitic steels. Materials Science and Engineering: A, 763, 138142.~~

Formatted: Indent: Left: 0", First line: 0"

Formatted: Normal

[58~~49~~] Bisong, E. (2019). More supervised machine learning techniques with scikit-learn. In Building Machine Learning and Deep Learning Models on Google Cloud Platform (pp. 287-308). Apress, Berkeley, CA.

[59~~0~~] Dong, X., Zhang, X., Du, K., Zhou, Y., Jin, T., & Ye, H. (2012). Microstructure of carbides at grain boundaries in nickel based superalloys. Journal of Materials Science & Technology, 28(11), 1031-1038.

[60~~51~~] Antonov, S., Chen, W., Huo, J., Feng, Q., Isheim, D., Seidman, D. N., ... & Tin, S. (2018). MC carbide characterization in high refractory content powder-processed Ni-based superalloys. Metallurgical and Materials Transactions A, 49(6), 2340-2351.

[61~~52~~] Mangen, W., & Nembach, E. (1989). The effect of grain size on the yield strength of the γ′-hardened superalloy NIMONIC PE16. Acta Metallurgica, 37(5), 1451-1463.

[62~~53~~] MacDonald, J. E., Khan, R. H. U., Aristizabal, M., Essa, K. E. A., Lunt, M. J., & Attallah, M. M. (2019). Influence of powder characteristics on the microstructure and mechanical properties of HIPped CM247LC Ni superalloy. Materials & Design, 174, 107796.

[63] Roth, H. A., Davis, C. L., & Thomson, R. C. (1997). Modeling solid solution strengthening in nickel alloys. Metallurgical and Materials Transactions A, 28(6), 1329-1335.

~~[54] Roth, H. A., Davis, C. L., & Thomson, R. C. (1997). Modeling solid solution strengthening in nickel alloys. Metallurgical and Materials Transactions A, 28(6), 1329-1335.~~

Formatted: Normal

Formatted: Normal

Formatted: Normal